REVIEW

DOI: https://doi.org/10.17816/socm622965

Problematic aspects of medical artificial intelligence. Part 2

Vitalii A. Berdutin[1], Tatyana E. Romanova[2], Sergey V. Romanov[3], Olga P. Abaeva[1]

[1] State Research Center — Burnasyan Federal Medical Biophysical Center, Moscow, Russia;
[2] Privolzhsky Research Medical University, Nizhny Novgorod, Russia;
[3] Privolzhsky District Medical Center, Nizhny Novgorod, Russia

ABSTRACT

The capabilities of artificial intelligence (AI) and machine learning are growing at an unprecedented pace. These technologies have many useful applications, from machine translation to medical image analysis.

A large number of such applications are currently being developed, and an increasing number of such applications is expected in the long term. Unfortunately, weaknesses and other unpleasant aspects of AI have received insufficient attention. In this review, we consider a whole range of already known problems and possible risks associated with the use of innovative neural network technologies, paying special attention to the ways of preventing real dangers and potential threats in order to expand the range of stakeholders and profile experts involved in the discussion of current issues of medical AI cybersecurity, formation of responsible approach to the vulnerabilities of neural network platforms, increasing the reliability of equipment protection for its safe use, as well as the importance of legal and ethical aspects of regulating the use of AI.

Despite certain challenges described in our review, it is clear that AI will be an important element of the healthcare future. As the population continues to age and the demand for healthcare services grows, neural networks are expected to drive healthcare very soon, especially in the areas of medical image analysis, virtual assistants, drug development, treatment recommendations and patients' data processing. We would like to emphasize that while we recognize the innovative role that digital technologies and AI can and should play in strengthening the domestic healthcare system, we should not overlook the importance of timely and accurate assessment of their beneficial or negative impact on the industry to ensure such management decisions do not unnecessarily divert our attention and resources from non-digital approaches and research.

This article is a continuation of the article by Berdutin VA, Romanova TE, Romanov SV, Abaeva OP. Problematic aspects of medical artificial intelligence. Part 1. Sociology of Medicine. 2023;22(2):202–211. DOI: https://doi.org/10.17816/socm619132

Keywords: artificial intelligence; machine learning; neural network.

To cite this article:
Berdutin VA, Romanova TE, Romanov SV, Abaeva OP. Problematic aspects of medical artificial intelligence. Part 2. *Sociology of Medicine.* 2024;23(1):94–103. DOI: https://doi.org/10.17816/socm622965

Submitted: 04.11.2023 **Accepted:** 10.09.2024 **Published online:** 13.09.2024

НАУЧНЫЙ ОБЗОР

DOI: https://doi.org/10.17816/socm622965

Проблемы медицинского искусственного интеллекта. Часть 2

В.А. Бердутин[1], Т.Е. Романова[2], С.В. Романов[3], О.П. Абаева[1]

[1] Государственный научный центр Российской Федерации — Федеральный медицинский биофизический центр имени А.И. Бурназяна, Москва, Россия;
[2] Приволжский исследовательский медицинский университет, Нижний Новгород, Россия;
[3] Приволжский окружной медицинский центр, Нижний Новгород, Россия

АННОТАЦИЯ

Возможности искусственного интеллекта (ИИ) и машинного обучения растут беспрецедентными темпами. Эти технологии имеют множество полезных применений: от машинного перевода до анализа медицинских изображений.

В настоящее время разрабатывается множество таких приложений, а в долгосрочной перспективе ожидается лавинообразное нарастание их числа. К сожалению, слабостям и иным неприятным сторонам ИИ уделяется недостаточно внимания. В данном обзоре мы рассматриваем целый спектр уже известных проблем и возможных рисков, связанных с использованием инновационных нейросетевых технологий, обращая особое внимание на способы предотвращения реальных опасностей и потенциальных угроз с целью расширить круг заинтересованных лиц и профильных экспертов, участвующих в обсуждении актуальных вопросов кибербезопасности медицинского ИИ, формирования ответственного подхода к уязвимостям нейросетевых платформ, повышения надёжности защиты оборудования для его безопасного использования, а также к важности правовых и этических аспектов регулирования применения ИИ.

Несмотря на отдельные проблемы, описанные в нашем обзоре, очевидно, что ИИ будет важным элементом будущего здравоохранения. Поскольку население продолжает стареть, а спрос на медицинские услуги растёт, ожидается, что нейронные сети совсем скоро будут выступать в роли движущей силы здравоохранения, особенно в областях анализа медицинских изображений, виртуальных помощников, разработки лекарств, рекомендаций по лечению и обработки данных пациентов. Мы хотели бы подчеркнуть, что, признавая инновационную роль, которую цифровые технологии и ИИ могут и должны играть в укреплении отечественной системы здравоохранения, не стоит упускать из виду, насколько важно своевременно и правильно оценивать их благоприятное или негативное влияние на отрасль, чтобы обеспечить такие управленческие решения, которые бы неоправданно не отвлекали наше внимание и ресурсы от нецифровых подходов и исследований.

Настоящая статья представляет собой продолжение статьи: Бердутин В.А., Романова Т.Е., Романов С.В., Абаева О.П. Проблемы медицинского искусственного интеллекта. Часть 1 // Социология медицины. 2023. Т. 22, № 2. С. 202–211. DOI: https://doi.org/10.17816/socm619132

Ключевые слова: искусственный интеллект; машинное обучение; нейронная сеть.

Как цитировать:
Бердутин В.А., Романова Т.Е., Романов С.В., Абаева О.П. Проблемы медицинского искусственного интеллекта. Часть 2 // Социология медицины. 2024. Т. 23, № 1. С. 94–103. DOI: https://doi.org/10.17816/socm622965

Рукопись получена: 04.11.2023 **Рукопись одобрена:** 10.09.2024 **Опубликована online:** 13.09.2024

ЭКО•ВЕКТОР

BACKGROUND

The healthcare industry and medical practice have undergone significant changes in recent years, driven by artificial intelligence (AI) technologies. AI platforms open up bright prospects that are becoming known to the medical community thanks to numerous scientific publications, as well as computer applications and gadgets being implemented everywhere. This publication continues our series of articles devoted to the promising and problematic aspects of using AI systems in healthcare, ranging from the problems of collecting personal information about a patient to the risks of using high-precision robotic surgeons. We are trying to draw the attention of the medical community to a number of weaknesses of AI, arising in connection with the use of new neural network platforms. Moreover, we highlight issues involved and describe the potential impacts and challenges to medical professionals and diagnosticians. Thus, the most pressing problems are unauthorized access to medical documents, which threatens with negative economic, psychological and reputational consequences, poorly structured, insufficient or falsified information, instability of remote management of medical devices and failures in them work, as well as the fundamental task of educating *Homo technicus* and many other problems. Over the past few decades, attention to the many ethical implications of AI has increased significantly. This includes the existential risk associated with further improvements in artificial general intelligence, which is a still hypothetical but extremely dangerous form of AI that is capable of much more intelligent actions than humans. This has led to extensive research into how humanity can avoid losing control to AI, which is far smarter than the best of us. The development of friendly AI is actively underway, which should be AI that is not hostile to people. Everyone's focus should be on the ethics of AI and the value of friendliness itself. In our publications, we briefly discuss a number of specific issues affecting the use of AI and machine learning (ML) in medicine, such as fairness, privacy, anonymity, interpretability, as well as some broader social issues such as ethics and legislation. We reckon that all of these are relevant aspects to consider in order to achieve the objective of fostering acceptance of AI and ML-based technologies, as well as to comply with an evolving legislation concerning the impact of digital technologies on ethically and privacy sensitive matters [1–6].

ABOUT THE FAILURES OF MEDICAL ARTIFICIAL INTELLIGENCE

Information technology professionals are well aware of the problem of AI platform performance degradation once it is deployed in the real world. For obvious reasons, developers try not to advertise minor failures and even major failures. However, public scandals with the giants of the IT industry are difficult to silence, and they have become public more than once. One of the biggest came when Google's Verily Health Sciences conducted field trials of its system for detecting diabetic retinopathy in Thailand. As the researchers described in an academic paper [7], the system performed poorly due to insufficient lighting and the presence of a large number of low-resolution images. 21% of the images that technicians attempted to input were rejected by the model as inappropriate. For the remaining images, the authors do not disclose accuracy figures, but say that performance has decreased noticeably. The system also often took a long time to get up and running because images had to be uploaded to the cloud, reducing the number of people the clinic could handle each day.

Skin cancer detection using a smartphone is one of the most promising applications of AI. However, every skin cancer detection system tested today cannot avoid making mistakes when it comes to non-white skin. A recent study quantified this for three commercial systems: ModelDerm, DeepDerm, HAM 10000. None of the systems performed better than a specialist physician, and all showed a significant decrease in performance between light and dark skin. For two systems, the sensitivity drop was about 50% in two sets of problems (0.41→0.12, 0.45→0.25, 0.69→0.23, 0.71→0.31). The third model actually showed worse sensitivity for lighter skin, but it also failed completely at the operating point the manufacturer used, achieving a sensitivity of <0.10 across the board. Additionally, dermatologists, who typically provide visual labels for AI training and testing datasets, were also found to perform worse on images of dark skin tones and unusual diseases compared to biopsy annotations [8].

Diagnosing breast cancer through mammography is probably the most studied application of computers in medical imaging, going back decades. In the mid-2010s, several Computer Aided Detection mammography software packages were released, which had numerous shortcomings and caused radiologists to still miss 16% of breast cancer cases. This could be an ideal application of AI capabilities, but despite intensive efforts in this direction for more than 20 years, the true level of an expert radiologist has not yet been achieved. Promising results in small studies are not replicated in larger studies. One of the most recent reviews, published in September 2021, reported that 34 out of 36 (94%) AI systems were less accurate than a single physician's judgment; and all framework models were less accurate than the consensus of two or more experts. AI systems lack the specificity to replace double reading of images by a radiologist in screening programs [9]. Needless to say, all the shortcomings of AI in radiology, arising both from government and private entities, undermine the medical community's trust in AI. But restoring lost trust may require a lot of time and effort.

The Epic Sepsis Model (ESM) has been implemented in hundreds of United States clinics to monitor patients and send alerts if they were at high risk for sepsis. The model uses a combination of real-time emergency department monitoring data: heart rate, blood pressure, etc., as well as

demographic information and information from the patient's medical records. In total, more than 60 functions are used. The diagnosis of sepsis was established by the model based on criteria from the recommendations of the Center for Disease Control and Prevention and International Classification of Diseases 10th Revision, in the presence of 2 criteria characteristic of systemic inflammation syndrome and 1 criterion of organ dysfunction, recorded within 6 hours. Model discrimination was assessed using the area under the receiver operating characteristic curve at the hospitalization level and with prediction horizons of 4, 8, 12, and 24 hours. External validation showed very poor model performance: area under curve 0.63 vs. reported areas under curve 0.73 and 0.83. Of the 2552 patients with sepsis, only 33% were identified, causing many false alarms. A proactive team of forensic scientists conducted an investigation into the ESM approach to sepsis prediction to show how shifts in data distribution can lead to ML model errors. It was found that changes in coding standards contributed to the decline in ESM performance over time, and that spurious correlations in model training data also played a negative role [10].

The most high-profile failure was International Business Machines (IBM) Watson Health division. The hype in the press was so great that rumors began to circulate about an "AI winter". Building on the success of its Watson system for Jeopardy, IBM launched Watson Health about 10 years ago to revolutionize healthcare with AI. It started with a highly touted partnership with Memorial Sloan Kettering to train AI on electronic health record data to make treatment recommendations. IBM chief executive officer J. Rometty called it "our moonshot". At its peak, Watson Health employed 7,000 people. However, IBM just recently sold off all of Watson Health piecemeal for about $1 billion. For comparison, IBM spent more than $5 billion to create Watson Health. IBM executives must have decided that the division had absolutely no chance of breaking even and decided to quickly liquidate it. For an uncompromising expose of the Watson Health collapse, visit https://slate.com. The central theme of the publication is: "When you try to combine high-tech bravado with a commitment to achieving stated goals in the health sector, you have to provide absolutely irrefutable evidence that you can achieve what you say.". Watson Health was expected to change healthcare in many important ways, providing information to oncologists about treating cancer patients, pharmaceutical companies about drug development, helping to conduct clinical trials, and more. This sounded revolutionary, but it never actually worked because IBM constantly needed huge amounts of data to train the model, which was simply impossible to find even for a lot of money. IBM partners, such as MD Anderson Cancer Center in Texas, pulled out one by one after participating physicians complained that the program did not have enough data to make the necessary recommendations [11].

NEURAL NETWORKS AND THEIR BIASES

Among the many concerns about AI that are attracting the attention of the medical community, the most controversial and, at the same time, pressing issue is the problem of identifying biases in AI algorithms. Previously, we have already briefly mentioned the troubles associated with the problem of systematic errors, which are caused by the lack of verified datasets, unidentified algorithms, incorrect classification, observational errors and illiterate software maintenance. Reasonable concerns about the development of biases in AI during operation have motivated the desire of developers to build fair, bias-free models, which is very laudable, but in reality is not easy. A fair AI model appears to be a bias-free predictive adaptive model. It should not be as if blocked from the outside world, i.e. should not "cook in its own juice". On the contrary, the neural network must continue to learn, constantly improving its performance, which will eventually lead to its implementation as a full-fledged electronic medical record administrator. Thus, the future hypothetical fair model will be able to independently function in a decision support mode, which, however, will not be autonomous, that is, doctors and patients will retain the right to make the final decision [12, 13].

At the same time, even such a seemingly maximally fair model may directly or indirectly have so-called hidden biases. Just as latent biases are typically described as errors waiting to happen, in complex software frameworks implicit bias refers to biases waiting to happen. It is common to identify three main problems associated with bias in AI algorithms.

1. The first major bias issue for this hypothetically fair algorithm is that, as an adaptive model, it may become biased over time. This can happen in several ways. An AI algorithm trained to perform fairly in one context can learn from differences in a healthcare organization's operational practices and begin generating biased results. Also, the neural network itself can learn from widespread, traditional and sometimes ridiculous prejudices found in the field of Russian healthcare, which one way or another lead to undesirable and even very unpleasant consequences.

For example, an algorithm for predicting patient mortality or an individual patient's response to certain treatment methods may well focus on existing ethnic or socio-economic differences in the living conditions of certain groups of patients and predict worse treatment results for them. But doing so can create a negative feedback loop whereby biases become stronger over time, further exacerbating the model's outlier prediction. From a clinical point of view, such a deviation is undesirable, since a correct forecast would make it possible to redirect healthcare resources precisely to bottlenecks and give correct recommendations for subsequent medical care, for example, to strengthen the palliative care system for socially vulnerable segments of the population. More importantly, it is now well known that generation of biases is quite possible even if the AI is prohibited from making inferences based on some variable, say the nationality or

address of a patient, when the data set does not include such a variable. Unfortunately, this can happen if other variables are correlated with or are proxies for the taboo variable, making the strategy of excluding variables of concern futile.

2. The next set of bias issues arise from the interaction of AI with the clinical environment, which includes its own implicit and explicit biases. It is worth noting 2 phenomena observed during the interaction between the patient and the doctor. One of them is the phenomenon of automation bias, in other words, an uncritical attitude towards AI recommendations that should be strictly obeyed. Even if an algorithm is used simply as a decision support tool, it can become a de facto autocrat if its orders are always followed. Overworked and time-constrained physicians who also avoid legal liability for ignoring algorithm recommendations may be oblivious to AI bias. Another is the phenomenon of privilege bias, which is the disproportionate advantage of individuals who already have privilege. Even the fairest algorithm can be unfair if it is used only in certain settings, for example, in private clinics serving mainly wealthy citizens. The class distrust of the precariat towards elite medical organizations, which will primarily use AI, may ultimately result in a general distrust of patients in its recommendations [14, 15].

3. The third type of bias, possible even in honest algorithms, is associated with the choice of the purpose of creating the model, its interest in a certain result. It is somewhat similar to the first type, but when the outcomes of interest or problems chosen to be solved by AI do not reflect the interests of individual patients or communities, it is essentially bias, namely the preferential selection or promotion of one outcome over another. with others. An illustration of what has been said will be the following. One of the reasons why many clinical trials have failed to improve the quality of care is the selection of surrogates for study outcomes that are not directly related to the fact of patient recovery. For example, treatment outcomes for heart failure were assessed only by changes in physiological parameters (left ventricular ejection fraction), and not by a decrease in disease symptoms: decreased fatigue and increased exercise tolerance. Results that are of interest to some stakeholder groups may not be of interest to others and vice versa. Patients are most concerned about restoring their own health and reducing treatment costs; they care little about the effectiveness of the health care system as such. Therefore, before introducing AI algorithms into daily clinical practice, we recommend risk management and bias prevention efforts.

AI decisions with high risks, for example, on chemotherapy treatment or artificial ventilation of the lungs, as well as decisions on the appointment of state social benefits, determination of disability, which are difficult to challenge, deserve especially close attention of the medical community. Today, we often encounter models that produce statements like: "Patients similar to you in a similar situation chose this and that.". Such an algorithm should be considered adaptively biased because its selection could obviously be influenced by outdated or inappropriately interpreted decision options. Unfortunately, we are not aware of any evidence that AI bias concerns have been addressed. But algorithmic biases that arise over time should be considered unfavorable events; in practice they mean that some patients may be harmed. Disparate AI behavior that is driven by bias and causes harm to patients should be subject to mandatory reporting on smart medical device performance. Since we are used to the fact that the doctor always controls when a drug is useful for some patients, but harmful for others, it is expected that a similar requirement should be presented to AI algorithms.

Aberrations in AI algorithms can arise not only from bias in the training data, but also from the way neural networks are trained over time and used in practice. Given the prevalence of prejudice, there is no excuse for being careless about it. Failure to proactively address biases, especially hidden biases that emerge unexpectedly, only exacerbates disparities among patient populations, undermines public trust in the healthcare system, and, paradoxically, ultimately hinders the accelerated adoption of medical AI [16].

DATA OF ANALYTICAL REPORTS OF INTERNATIONAL ORGANIZATIONS ON THE CHALLENGES ASSOCIATED WITH MEDICAL ARTIFICIAL INTELLIGENCE TECHNOLOGIES

With increased interest in digital health, there is a large number of AI adoptions without carefully examining the evidence base for benefits and harms. Excessive enthusiasm for digitalization has led to a proliferation of short-lived implementations and a huge variety of digital tools with limited understanding of their impact on the health system and people's well-being. World Health Organization experts on this matter stated: "To improve health and reduce health inequalities, careful evaluation of eHealth is needed to generate evidence and promote appropriate integration and use of technologies." [17].

Recently, several interesting analytical reports have been released that touch on the topic of AI in healthcare. Here are some excerpts from them. KLAS Research and the College of Healthcare Information Management Executives published the report "Healthcare AI 2019. Actualizing the potential of artificial intelligence" about the first real cases of integrating AI systems into practical medicine, which related to predicting re-hospitalizations and reducing unnecessary emergency calls. KLAS and College of Healthcare Information Management Executives surveyed 57 healthcare organizations that had recently implemented ML and natural language processing systems to assess clinical, financial and operational advancements. KLAS assessed customer satisfaction for six leading healthcare AI providers: Jvion, DataRobot, KenSci, Clinithink, IBM Watson Health and Health Catalyst. Among other things, the report

focused special attention on the failures of IBM Watson; the authors of the study came to the conclusion that IBM failed to correct the situation with its product [18].

OptumIQ published the "Annual Survey on AI in Health Care" report, which analyzed a survey of 500 healthcare executives and concluded that the number of AI implementations in medicine increased by almost 88% compared to the previous year. The authors note the skepticism of some reputable healthcare organizations regarding the further growth of investments in AI, since they are not at all confident that the costs incurred will pay off at least within a three-year period [19]. CB Insights published a report showing that while investor interest in AI for healthcare surged in 2019, the investment climate is likely to cool somewhat going forward. The wake-up call came from Freenome, a company using neural networks for early cancer detection, which closed a $160 million Series B round of funding in July 2019 [20].

The American Hospital Association's Center for Healthcare Innovation released a report "AI and Care Delivery: Emerging opportunities for AI to transform how care is delivered". It explores the use of AI as a clinical decision support tool, based on the opinions of healthcare experts. The report, in particular, examines ways to solve numerous problematic issues, including reducing the enormous costs of AI throughout the entire cycle of care. Massachusetts Institute of Technology Technology Review released a report "The AI Effect. How artificial intelligence is making healthcare more human". It features data from a survey of more than 900 healthcare professionals conducted by Massachusetts Institute of Technology Technology Review Insights in collaboration with General Electric Healthcare. The survey found that only 72% of respondents expressed direct interest in implementing AI. 20% of respondents believe that AI will not be able to improve their economic situation, and 19% that AI will not be able to make a medical institution more competitive and customer-oriented. As investment in medical AI picks up pace, respondents with existing deep learning projects are worried about spending more and more money on algorithm maintenance every year. These findings are significant for the industry as health care delivery and management become increasingly complex and costly, and professional and technological capacity becomes increasingly burdensome. Given that doctors are stuck in the routine of an ever-increasing workload and stupid, low-paid work, their partial replacement with chatbots will finally deprive patients of live interaction with doctors [21].

KPMG has released a report "Healthcare insiders: Taking the temperature of artificial intelligence in healthcare". It confirms the growing interest in the use of AI in medicine. However, the negative point is that 32% of respondents do not see the prospect of AI for objectively assessing the condition of patients. The main barriers to AI are the lack of qualified personnel, high costs of creating AI systems and high risks of privacy violations [22].

PROBLEM ASPECTS OF USING VARIOUS MACHINE LEARNING ALGORITHMS

K-Nearest Neighbors (KNN), Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) are considered the most popular ML algorithms. They differ in their approaches, strengths and weaknesses, and use cases. To be brief, the difference between these algorithms is as follows.

1. KNN is a simple and universal algorithm used for both classification and regression problems. It works on the principle of finding the k-nearest data points to a given query point based on a distance metric. In classification, KNN assigns the majority class among k-nearest neighbors as the predicted class for the query point. In regression, KNN takes the average or weighted average of the target k-nearest neighbor values as the predicted value for the query point. KNN is nonparametric, meaning it makes no assumptions about the underlying distribution of the data. This requires a lot of computing power, especially for large data sets, because distance calculations must be made for all data points.

2. Random Forest is an ensemble learning method based on decision trees and is mainly used for classification and regression problems. During training, it creates multiple decision trees, where each tree is trained on a random subset of the features and seed data. In classification, the final prediction is based on the majority votes of the individual trees. Regression requires the average prediction of individual trees. RF mitigates overfitting and achieves good generalization by combining predictions from multiple trees. It handles multi-dimensional data well and is less prone to outliers.

3. XGBoost is an advanced gradient boosting algorithm used for classification, regression, and ranking problems. Like RF, it also works with an ensemble of decision trees, but builds the trees sequentially rather than independently. XGBoost uses a gradient boosting system to optimize the ensemble by minimizing the loss function. It uses regularization techniques to avoid overfitting and improve model performance. XGBoost is computationally efficient and can process large data sets efficiently. It often outperforms other algorithms in various ML competitions and real-world applications.

The application of the listed ML models depends on the specific task for which they are going to be used, for example, regression or classification. Generally speaking, the following features of these models need to be taken into account.

K-Nearest Neighbors is simple and intuitive, applicable to both classification and regression problems. In KNN, the prediction for a new data point is based on the majority class (for classification) or the average of its K-nearest neighbors (for regression) in the feature space. The K value is a hyperparameter that determines how many neighboring points should be considered.

Advantages of the algorithm:

- Easy to understand and implement.
- Nonparametric, meaning no assumptions are made about the underlying distribution of the data.
- Works well on small data sets with simple decision boundaries.

Flaws:
- Can be computationally expensive for large data sets because it requires calculating distances to all data points.
- Sensitive to non-essential functions and noise.
- Does not handle imbalanced data sets well.

RF is an ensemble learning method that combines multiple decision trees to produce more accurate predictions. During training, it builds multiple decision trees and averages their predictions to improve reliability and accuracy. Each tree is trained on a random subset of data and a random subset of features, which mitigates overfitting and increases generalization.

Advantages:
- Robust to overfitting and works well with a wide range of data types.
- Handles multi-dimensional data well.
- Can provide feature importance ratings.

Flaws:
- Can be slow to train and predict large data sets.
- Lacks transparency and interpretability compared to standalone decision trees.
- May not perform as well as more advanced models like XGBoost for some complex tasks.

XGBoost is an advanced implementation of gradient boosting that is an ensemble technique that combines weak decision trees to create a strong predictive model. XGBoost improves on traditional gradient boosting by incorporating regularization conditions, parallel processing, and efficient data manipulation to achieve higher accuracy and speed.

Advantages:
- High predictive efficiency due to the boosting mechanism
- Handles poorly representative data well.
- Supports regularization to prevent overfitting.
- Fast and scalable thanks to parallelization.

Flaws:
- Requires hyperparameter tuning, which can be time consuming.
- More complex than basic models such as KNN and Random Forest.
- Tends to overtrain if not well-tuned.

So, let's summarize the review of algorithms. KNN is a simple and interpretable algorithm suitable for small data sets, and Random Forest is a powerful ensemble method that provides robust performance and feature importance. XGBoost is an advanced precision boosting algorithm that is highly accurate and suitable for large-scale data sets. The choice of model depends on the specific characteristics of the input data, the size of the data sets, and the desired balance between simplicity and forecasting performance [23, 24].

CONCLUSION

Although hundreds of AI algorithms have received approval from government health regulators around the world, such as the United States Food and Drugs Administration, neural network platforms are prone to implicit bias and inconsistent generalizations, especially if the analyzed data is insufficient or incorrect. There remains a glimmer of hope that generative AI could reduce the need for real data, but its usefulness remains unclear. Dermatological diseases serve as a very illustrative example of synthetic image generation due to the variety of pathological manifestations, especially taking into account the color and tone of the patient's skin. Scalable latent diffusion algorithms can generate images of skin diseases to further train the model, which can certainly improve its performance in data-limited settings. However, performance gains are achieved when the ratio of synthetic to real images is more than 10:1; it is significantly less than the gain obtained from adding real images, so collecting objective data remains the main condition for ensuring the reliability of medical AI.

Medical AI is a potentially powerful tool, but its operation poses many challenges. To intelligently and successfully use this advanced, but still imperfect technology, without letting the genie out of the bottle, we need effective strategies and thoughtful management. This will require the training of medical personnel at a completely new level, who will be able to actively and methodically participate in the development, testing and use of highly complex innovative neural network models. This, in turn, will require a fundamental overhaul and complete renewal of programs for training and certification of digital health professionals, including the next generation of future digital health professionals who can ensure the safety of AI in the clinical environment. Such steps will be necessary to maintain public trust in medicine in the coming era of AI.

Despite all the challenges described in some of our reviews, it is clear that AI will be an important part of the future of healthcare. As the population continues to age and the demand for healthcare services increases, neural networks are expected to play a critical role in healthcare, especially in the areas of medical image analysis, virtual assistants, drug development, medical treatment recommendations, and patient data processing. Advanced AI algorithms will be able to analyze computed tomography, magnetic resonance imaging, positron emission tomography — computed tomography images with a level of accuracy comparable to or even greater than that of radiologists. All this in general can help doctors more accurately diagnose diseases and quickly assess the conditions of patients, which will lead to improved quality and accessibility of medical care in the country. However, to achieve such ambitious goals, we will need truly pragmatic actions and a responsible attitude towards AI technologies. Legislation governing the use of AI and ML algorithms should explicitly include reference to tracking performance variations, including those that occur during operation.

In conclusion, we would like to emphasize that, while recognizing the innovative role that digital technologies and AI can and should play in strengthening the Russian healthcare system, we must not lose sight of how important it is to timely and correctly assess their enabling or negative impact on the industry in order to ensure such management decisions that do not unduly divert resources from alternative, non-digital approaches.

ADDITIONAL INFORMATION

This article is a continuation of the article by Berdutin VA, Romanova TE, Romanov SV, Abaeva OP. Problematic aspects of medical artificial intelligence. Part 1. Sociology of Medicine. 2023;22(2):202–211. DOI: https://doi.org/10.17816/socm619132

Competing interests. The authors declare that they have no competing interests.

Funding source. Not specified.

Author's contribution. All authors confirm compliance of their authorship with the international ICMJE criteria. The largest contribution is distributed as follows: all authors — review concept, collection and processing of materials; V.A. Berdutin — writing the text; T.E. Romanova — editing.

REFERENCES

1. Reshetnikov AV, Shamshurina NG, Shamshurin VI. *Economics and management in healthcare: Textbook and workshop.* 2nd ed. Moscow: Yurayt Publishing House; 2020 (In Russ.) EDN: KSZBPT

2. Reshetnikov A, Fedorova J, Prisyazhnaya N, et al. Health management for sustainable development. In: *2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*. IEEE, 2018.

3. Berdutin V. *Socionic vision on Bioethics and Deontology.* LAP LAMBERT Academic Publishing; 2018.

4. Liu J. Artificial Intelligence and Data Analytics Applications in Healthcare General. Review and Case Studies. In: *CAIH2020: Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare, October 2020*, P. 49–53. doi: 10.1145/3433996.3434006

5. Daley K. Two arguments against human-friendly AI. *AI and Ethics.* 2021;1(4):435–444. doi: 10.1007/s43681-021-00051-6

6. Vellido A. Societal Issues Concerning the Application of Artificial Intelligence in Medicine. Kidney Dis. 2019;5(1):11–17. doi: 10.1159/000492428

7. Breede E, Bayor E, Hersh F, et al. *A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy.* In: CHI 2020; 2020 Apr 25–30; Honolulu. P. 1–12. doi: 10.1145/3313831.3376718

8. Daneshjou R, Vodrahalli K, Novoa RA, et al. *Disparities in Dermatology AI Performance on a Diverse, Curated Clinical Image Set* [Internet]. Cornell University; 2022. [cited 2023 Sep 05]. Available from: https://arxiv.org/ftp/arxiv/papers/2203/2203.08807.pdf

9. Freeman K, Geppert J, Stinton Ch, et al. Use of artificial intelligence for image analysis in breast cancer screening programs: systematic review of test accuracy. *BMJ.* 2021;374:n1872. doi: 10.1136/bmj.n1872

10. Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med.* 2021;181(8):1065–1070. doi: 10.1001/jamainternmed.2021.2626

11. O'Leary L. How IBM's Watson Went from the Future of Health Care to Sold Off for Parts. In: *Slate* [Internet]. 2022. [cited 2023 Sep 23]. Available from: https://slate.com/technology/2022/01/ibm-watson-health-failure-artificial-intelligence.html

12. Khan B, Hajira F, Qureshi A, et al. Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector. In: *Biomedical Materials & Devices*; 2023. Feb 8. P. 1–8. doi: 10.1007/s44174-023-00063-2

13. Lee TT, Kesselheim AS. U.S. Food and Drug Administration Precertification Pilot Program for Digital Health Software: Weighing the Benefits and Risks. *Ann Intern Med.* 2018;168(10):730–732. doi: 10.7326/M17-2715

14. Parikh RB, Teeple S, Navathe AS. Addressing Bias in Artificial Intelligence in Health Care. *JAMA.* 2019;322(24):2377–2378. doi: 10.1001/jama.2019.18058

15. Challen R, Denny J, Pitt M, et al. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf.* 2019;28(3):231–237. doi: 10.1136/bmjqs-2018-008370

16. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med.* 2019;25(1):30–36. doi: 10.1038/s41591-018-0307-0

17. *Monitoring the implementation of digital health:an overview of selected national and international methodologies* [Internet]. Copenhagen: WHO Regional Office for Europe; 2022. [cited 2023 Sep 20]. Available from: https://www.who.int/europe/publications/i/item/WHO-EURO-2022-5985-45750-65816

18. Gale A. Reimagined Hospitals. How Far Is the Future? *HealthManagement.org The Journal.* 2020;20(1):36–38.

19. Christensen J. A Snapshot of Imaging Technology: Exciting Developments and When to Expect Them. *HealthManagement.org The Journal.* 2020;20(6):476–479.

20. Landi H. Investors poured $4B into healthcare AI startups in 2019. In: *Fierce Healthcare* [Internet]. Questex; 2020 [cited 2023 Sep 23]. Available from: https://www.fiercehealthcare.com/tech/investors-poured-4b-into-healthcare-ai-startups-2019

21. Memora Health raises $40M for its virtual care delivery platform. Memora Health competitors include Wheel, Welby Health, and Twistle. ResearchBriefs. In: *CBinsights* [Internet]; 2022. [cited 2023 Sep 24] Available from: https://www.cbinsights.com/research/memora-health-competitors-wheel-welby-health-twistle/

22. The AI effect: How artificial intelligence is making health care more human. In: *Technology review* [Internet]. GE Healthcare. [cited 2023 Sep 13]. Available from: https://www.technologyreview.com/hub/ai-effect/

23. Avuçlu E. Determining the most accurate machine learning algorithms for medical diagnosis using the monk' problems database and statistical measurements. *Journal of Experimental & Theoretical Artificial Intelligence.* Forthcoming. 2023. doi: 10.1080/0952813X.2023.2196984

24. Shukla S. Enhancing Healthcare Insights, Exploring Diverse Use-Cases with K-means Clustering. *International Journal of Management, IT & Engineering.* 2023;13(8):60–68.

СПИСОК ЛИТЕРАТУРЫ

1. Решетников А.В., Шамшурина Н.Г., Шамшурин В.И. Экономика и управление в здравоохранении. 2-е изд. Москва: Издательство Юрайт, 2020. EDN: KSZBPT
2. Reshetnikov A., Fedorova J., Prisyazhnaya N., et al. Health management for sustainable development. В кн.: 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4). IEEE, 2018.
3. Berdutin V. Socionic vision on Bioethics and Deontology. Lap Lambert Academic Publishing, 2018.
4. Liu J. Artificial Intelligence and Data Analytics Applications in Healthcare General. Review and Case Studies. In: CAIH2020: Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare; Oct 2020. P. 49–53. doi: 10.1145/3433996.3434006
5. Daley K. Two arguments against human-friendly AI // AI and Ethics. 2021. Vol. 1, N 4. P. 435–444. doi: 10.1007/s43681-021-00051-6
6. Vellido A. Societal Issues Concerning the Application of Artificial Intelligence in Medicine // Kidney Dis. 2019. Vol. 5, N 1. P. 11–17. doi: 10.1159/000492428
7. Breede E., Bayor E., Hersh F., et al. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. In: CHI 2020; 2020 Apr 25–30; Honolulu. P. 1–12. doi: 10.1145/3313831.3376718
8. Daneshjou R., Vodrahalli K., Novoa R.A., et al. Disparities in Dermatology AI Performance on a Diverse, Curated Clinical Image Set [Internet]. Cornell University, 2022. Режим доступа: https://arxiv.org/ftp/arxiv/papers/2203/2203.08807.pdf Дата обращения: 05.09.2023.
9. Freeman K., Geppert J., Stinton Ch., Todkill D., et al. Use of artificial intelligence for image analysis in breast cancer screening programs: systematic review of test accuracy // BMJ. 2021. Vol. 374. P. n1872. doi: 10.1136/bmj.n1872
10. Wong A., Otles E., Donnelly J.P., et al. External Validation of a Widely Implemented Sepsis Prediction Model in Hospitalized Patients // JAMA Intern Med. 2021. Vol. 181, N 8. P. 1065–1070. doi: 10.1001/jamainternmed.2021.2626
11. O'Leary L. How IBM's Watson Went from the Future of Health Care to Sold Off for Parts. B: Slate [интернет]. 2022. Режим доступа: https://slate.com/technology/2022/01/ibm-watson-health-failure-artificial-intelligence.html Дата обращения: 23.09.2023
12. Khan B., Hajira F., Qureshi A., et al. Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector. In: Biomedical Materials & Devices, 2023. Feb 8. P. 1–8. doi: 10.1007/s44174-023-00063-2
13. Lee T.T., Kesselheim A.S. U.S. Food and Drug Administration Precertification Pilot Program for Digital Health Software: Weighing the Benefits and Risks // Ann Intern Med. 2018. Vol. 168, N 10. P. 730–732. doi: 10.7326/M17-2715
14. Parikh R.B., Teeple S., Navathe A.S. Addressing bias in artificial intelligence in health care // JAMA. 2019. Vol. 322, N 24. P. 2377–2378. doi: 10.1001/jama.2019.18058
15. Challen R., Denny J., Pitt M., et al. Artificial intelligence, bias and clinical safety // BMJ Qual Saf. 2019. Vol. 28, N 3. P. 231–237. doi: 10.1136/bmjqs-2018-008370
16. He J., Baxter S.L., Xu J., et al. The practical implementation of artificial intelligence technologies in medicine // Nat. Med. 2019. Vol. 25, N 1. P. 30–36. doi: 10.1038/s41591-018-0307-0
17. Monitoring the implementation of digital health: an overview of selected national and international methodologies [Internet]. Copenhagen: WHO Regional Office for Europe, 2022. Режим доступа: https://www.who.int/europe/publications/i/item/WHO-EURO-2022-5985-45750-65816 Дата обращения: 20.09.2023.
18. Gale A. Reimagined Hospitals. How Far Is the Future? // HealthManagement.org The Journal. 2020. Vol. 20, N 1. P. 36–38.
19. Christensen J. A Snapshot of Imaging Technology: Exciting Developments and When to Expect Them // HealthManagement.org The Journal. 2020. Vol. 20, N 6. P. 476–479
20. Landi H. Investors poured $4B into healthcare AI startups in 2019. B: Fierce Healthcare [интернет]. Questex, 2020. Режим доступа: https://www.fiercehealthcare.com/tech/investors-poured-4b-into-healthcare-ai-startups-2019 Дата обращения: 23.09.2023
21. Memora Health raises $40M for its virtual care delivery platform. Memora Health competitors include Wheel, Welby Health, and Twistle. ResearchBriefs. B: CBinsights [интернет]. 2022. Режим доступа: https://www.cbinsights.com/research/memora-health-competitors-wheel-welby-health-twistle/ Дата обращения: 24.09.2023
22. The AI effect: How artificial intelligence is making health care more human. B: Technology review [интернет]. GE Healthcare. Режим доступа: https://www.technologyreview.com/hub/ai-effect/ Дата обращения: 13.09.2023
23. Avuçlu E. Determining the most accurate machine learning algorithms for medical diagnosis using the monk' problems database and statistical measurements. Journal of Experimental & Theoretical Artificial Intelligence. Forthcoming. 2023. doi: 10.1080/0952813X.2023.2196984
24. Shukla S. Enhancing healthcare insights, exploring diverse use-cases with K-means clustering // International Journal of Management, IT & Engineering. 2023. Vol. 13, N 8. P. 60–68.

AUTHORS' INFO

*** Vitalii A. Berdutin,** MD, Cand. Sci. (Medicine);
address: 46 Zhivopisnaya street, 123098 Moscow, Russia;
ORCID: 0000-0003-3211-0899;
eLibrary SPIN: 8316-7111;
e-mail: vberdt@gmail.com
Tatyana E. Romanova, MD, Cand. Sci. (Medicine);
ORCID: 0000-0001-6328-079X;
eLibrary SPIN: 4943-6121;
e-mail: drmedromanova@gmail.com

ОБ АВТОРАХ

*** Бердутин Виталий Анатольевич,** канд. мед. наук;
адрес: Россия, 123098, Москва, ул. Живописная, д. 46;
ORCID: 0000-0003-3211-0899;
eLibrary SPIN: 8316-7111;
e-mail: vberdt@gmail.com
Романова Татьяна Евгеньевна, канд. мед. наук;
ORCID: 0000-0001-6328-079X;
eLibrary SPIN: 4943-6121;
e-mail: drmedromanova@gmail.com

Sergey V. Romanov, MD, Dr. Sci. (Medicine);
ORCID: 0000-0002-1815-5436;
eLibrary SPIN: 9014-6344;
e-mail: director@pomc.ru
Olga P. Abaeva, MD, Dr. Sci. (Medicine), Professor;
ORCID: 0000-0001-7403-7744;
eLibrary SPIN: 5602-2435;
e-mail: abaevaop@inbox.ru

Романов Сергей Владимирович, д-р мед. наук;
ORCID: 0000-0002-1815-5436;
eLibrary SPIN: 9014-6344;
e-mail: director@pomc.ru
Абаева Ольга Петровна, д-р мед. наук, проф.;
ORCID: 0000-0001-7403-7744;
eLibrary SPIN: 5602-2435;
e-mail: abaevaop@inbox.ru

* Corresponding author / Автор, ответственный за переписку